The Impact of Aggregators on Internet News Consumption*

Susan Athey
Stanford Business School and NBER

Markus Mobius
Microsoft Research, University of Michigan and NBER

Jeno Pal
Central European University

January 11, 2017

Abstract

A policy debate centers around the question whether news aggregators such as Google News decrease or increase traffic to online news sites. One side of the debate, typically espoused by publishers, views aggregators as *substitutes* for traditional news consumption because aggregators' landing pages provide snippets of news stories and therefore reduce the incentive to click on the linked articles. Defendants of aggregators, on the other hand, view aggregators as *complements* because they make it easier to discover news and therefore drive traffic to publishers. This debate has received particular attention in the European Union where two countries, Germany and Spain, enacted copyright reforms that allow newspapers to charge aggregators for linking to news snippets. In this paper, we use Spain as a natural experiment because Google News shut down all together in response to the reform in December 2014. We compare the news consumption of a large number of Google News users with a synthetic control group of similar non-Google News users. We find that the shutdown of Google News reduces overall news consumption by about 20% for treatment users, and it reduces page views on publishers other than Google News by 10%. This decrease is concentrated

*The authors acknowledge support from Microsoft Research. We would like to thank seminar participants at Microsoft Research, University of Michigan, Stanford University, University of Chicago, the Toulouse Network for Information Technology, the WZB Lectures in Berlin, University of Mannheim, Columbia, Boston University, EARIE, IC2S2, Toulouse, Harvard Digital Business Seminar, and a variety of industry forums for helpful comments, as well as Josh Feng, Samuel Grondahl, Sebastian Steffen and Aaron Kaye for exceptional research assistance. This paper replaces our previous paper, "The Impact of News Aggregators on Internet News Consumption: The Case of Localization," which used a different natural experiment (in France) to address a similar question. We are currently updating the analysis for the French natural experiment to be consistent with the analytic approach introduced in this paper, and plan to report the updated results as a supplement to this paper.

around small publishers while large publishers do not see significant changes in their overall traffic. We further find that when Google News shuts down, its users are able to replace some but not all of the types of news they previously read. Post-shutdown, they read less breaking news, hard news, and news that is not well covered on their favorite news publishers. These news categories explain most of the overall reduction in news consumption, and shed light on the mechanisms through which aggregators interact with traditional publishers.

1 Introduction

A recent policy debate concerns the impact of the internet on the news media. Many authors have noted a series of stylized facts about the industry that suggest the impact of the internet has been quite negative: for example, the Newspaper Association of America reports that from 2000 to 2009, newspaper advertising revenue declined by 57% in real terms, and circulation fell by 18%. Digital has become quite important for news publishers: Pew Research reports that by 2015, a quarter of newspaper advertising revenue comes from digital, but digital revenue has not replaced lost revenue from traditional advertising. In addition to many widely publicized bankruptcies, investment in journalism has been reported to decline; for example, newsroom employment declined 40% between 1994 and 2014.[1] The issues cut across large and small publishers: a 2015 survey of U.S. digital publishers focusing on local news found that fewer than half were profitable.[2] At the same time, there has been popular discussion about how publishers and individual journalists have responded to the incentives created by the digital environment, for example, by optimizing the writing and headlines for search engines, aggregators, and social media.[3]

One particularly contentious point in this debate is the role of news aggregators. Pure aggregators such as Google News do not produce any original content but rather curate content created by other outlets through using a combination of human editorial judgement and computer algorithms. The results are presented with a few sentences and perhaps photos from the original article; to read the full article, users can click through and go to the web site of the original content creator. Thus, news aggregators act in dual roles: their front pages look very similar to news outlets who produce original content, and thus may be a substitute for them; yet they also aggregate a wide range of sources, and may be an effective mechanism for search and discovery, which places it in the role of an upstream complement to the outlets who produce news. The magnitudes of these two effects, as well as the answer to more detailed questions about how aggregators affect different types of outlets and readership of different types of news, determine whether aggregators increase or decrease the returns to investment in news reporting.

This debate has received particular attention in the European Union where two countries,

[1] See Barthel (June 15, 2016; accessed November 17, 2016).

[2] See Lu and Holcomb (June 15, 2016; accessed November 17, 2016)

[3] Numerous how-to guides and advice for reporters to modify articles and headlines have become over time; for an older one, see Smith (April 4, 2008; accessed November 17, 2016), while a more recent discussion can be found at Jafri (January 27, 2014; accessed November 17, 2016).

Germany and Spain, enacted copyright reforms that allow newspapers to charge aggregators for linking to news snippets. The German law came into effect of August 1, 2013 and allowed newspapers to provide a free license to aggregators. Members of the main German newspaper trade association VG Media provided Google News with a free license - hence, the introduction of the law had no impact on Google News in Germany. However, no other aggregator received such a free license and since August 2014 smaller German aggregators such as GMX, Web.de and T-Online have scaled down or discontinued their services. The Spanish law came into effect in January 2015 and did not provide for a free license – Google News therefore decided to shut down its news aggregator on December 16, 2014 and other aggregators such as Yahoo and Bing News have followed suit.

In this paper, we use the shutdown of Google News in Spain as a natural experiment to evaluate how news aggregators affect news consumption. Our dataset is a sample of all browsing events for more than 100,000 users in Spain who use Microsoft products for browsing the internet and have opted in to allowing their data to be used for research purposes.[4] We use this dataset to construct quasi-experimental treatment and control groups. Our treatment users are all Google News users. We match these users with a synthetic control group of non-Google News users who have the same news consumption patterns as the corresponding Google News users after the shutdown of Google News. Our matching procedure therefore selects control users who make the same consumption decisions in the absence of Google News (when both groups have access to the same news discovery technology) and therefore have the same underlying preferences.

We then first compare the *overall* news consumption of treatment and control users pre-shutdown across all news categories. This allows us to evaluate the impact on total traffic to newspapers if Google News is shut down – this quantity is of fundamental interest to publishers and policy makers who want to know whether aggregators steal traffic from news creators (substitutes view) or increase traffic (complements view). We find that the removal of Google News reduces overall news consumption (including consumption of the Google News homepage) by about 20% for treatment users, while visits to news publishers decline by about 10%. This decrease is concentrated around small publishers, while large publishers do not see significant changes in their overall traffic (but see an increase in their own home page views, offset by a decrease in views of articles).

These results highlight the potential impact of intermediaries on industry structure: they make it easier for consumers to search and consume products from small firms, increasing competition across publishers for consumer attention. We have seen similar effects in other technology-enabled intermediaries, such as eBay, Uber, AirBnb, and travel and price comparison sites, where the technology platform reduces search costs and enables smaller firms that may lack name recognition or reputation to be discovered by consumers. Whether this is good or bad for consumer welfare depends on details of how investment is spread across these firms, as well as how investment impacts market share. In the case of news, the welfare effects depend on whether the investments that increase visibility on aggregators are

[4]The data is subject to stringent privacy restrictions and at all times resides only on secure servers, and only aggregate statistics and the output of statistical models can be reported. However, we are able to construct the variables for analysis using the fully disaggregated data.

welfare-enhancing (unique investigative journalism) or wasteful (misleading headlines), as well as whether smaller firms add to the diversity of alternative perspectives rather than reproduce news where the investments have been made by others.

We further find that when Google News shuts down, its users are able to replace some but not all of the types of news they previously read. Post-shutdown, they read less breaking news, hard news, and news that is not well covered on their favorite news publishers. These news categories explain most of the overall reduction in news consumption. The result about breaking news highlights an advantage for aggregators: they can always be "first to market" with the latest news, since they can link to articles as soon as they appear on publisher home pages. They can also offer more breadth than any individual publisher, allowing readers to access topics not covered by their favorite outlets, as well as allowing readers to read more in-depth coverage on particular topics. Finally, the Google News homepage focuses more on hard news than the typical publisher home page.

Despite the intrinsic policy importance of the news industry and the close attention this issue has received from regulators, there is very little existing empirical evidence on the impact of aggregators. The paper closest to this one is Calzada and Gil (2016), which independently studies the same event using a different data source. Their paper finds an almost identical magnitude (an 11% reduction in visits to news outlets due to the Google News shutdown). Their paper uses aggregate data about site visits, while our paper relies on individual-level browsing data. As such, we are able to explore how individual consumers substitute the missing Google news consumption, and how the content of their consumption changes. The difference in data sources also affects the identification strategy: we are able to use Spanish users who were not previously Google News users as a control group, while Calzada and Gil (2016) relies on France and Germany to control for time trends in news viewing. A limitation of our study is that it only includes users of Microsoft products, who account for less than half of PC news browsing; thus, Calzada and Gil (2016) provides confirmation that their behavior is broadly representative.

Another closely related paper is Chiou and Tucker (2015). They study a "natural experiment" where Google News had a dispute with the Associated Press, and as a result, did not show Associated Press content for about seven weeks. The paper has aggregate data about page views to Google News as well as the sites visited immediately after Google news. They use views to Yahoo! News as a control. The paper finds that Google News is a complement to news outlets: taking the Associated Press content away from Google News leads to fewer visits to news outlets (where Associated Press articles are featured). Our paper is complementary to theirs: our main finding (complementarity) is consistent with theirs – however, our individual level data allows us to (a) dis-aggregate the effects by outlet size and (b) analyze the types of news consumption that see the biggest drops, which allows us a more nuanced analysis. Though less directly related to our work, a literature on the network structure of information flows on the web finds that "hubs" may improve information flows.[5]

[5]The prevailing vein of this work was pioneered by Kleinberg (1999), who developed the hub-authority information flow model and the Hyperlink-Induced Topic Search (HITS) algorithm (see also Kleinberg and Lawrence, 2001). Weber and Monge (2011) introduce a third category of nodes, sources, to Kleinberg's

The balance of the paper is organized as follows. In Section 2 we introduce a simple empirical model of news consumption and describe a matching algorithm that allows us to construct synthetic control and treatment groups. We then use this framework in Section 3 to document the drop in overall news consumption after the shutdown of Google News in Spain on December 16, 2014. We decompose this overall volume drop in Section 4 and show that it is predominantly driven by a reduction in the consumption of scarce and breaking news.

2 Empirical Model and Data Description

2.1 Theory

In this section we present a stylized model of news consumption that helps motivate our empirical strategy for estimating the effects of aggregators.

Preferences for News

A consumer i has one unit of time that she can allocate between reading news stories and other activities. Every news story has a vector c of characteristics, referred to as its "type," with components indexed by $d = 1, .., D$. For example, a particular characteristic might indicate whether a news story is breaking news or not, or whether the story is about a popular topic. The characteristics space is a finite product set $\mathcal{C} = \prod_{d=1,..,D} \mathcal{C}_d$. We denote user i's consumption of type c news at time t by $N_{i,c}^t$, and the vector of overall consumption is denoted N_i^t.

The total consumption of other leisure activities is denoted with L_i^t. The user's utility from consuming the bundle (N_i^t, L_i^t) at time t is described by a nested Cobb-Douglas utility function:

$$U_{it} = \left[\prod_{c \in \mathcal{C}} \left(N_{i,c}^t \right)^{\alpha_{i,c}} \right]^{\tau^t \tau^i} L^{1-\tau^t \tau^i} \tag{1}$$

The key implications of this functional form are described as follows. First, a user's share of time spent reading news can be decomposed into a date effect and a person effect: τ^t captures weekday and seasonal effects in preferences for reading news, while τ_i captures the importance that individual i attaches to news reading. Second, a utility maximizing individual will consume a constant share of news of a particular type, so long as the costs of finding different types of news do not change. We will formally check the assumption that preferences are stable over time below.

Discovery of News

Consumers do not directly know what articles are available to read on a given day. In order to discover articles, they must visit home pages of news outlets, view social media, use search,

model and use their expanded model to study news information flow. Using random graph models to study hyperlink structure, they find that hubs very rarely exist in pure form. Rather, there is often a degree of reciprocity such that hubs function to some degree as distribution nodes for information through the network.

or use aggregators.

Home pages or "landing pages" of news outlets merit special discussion. These pages have a list of article headlines and snippets from the stories, as well as links to the stories. Home pages play a dual role for consumers: they are a mechanism to learn about articles that are available, and users also consume news directly by reading the headline and the beginning of a news story. For simplicity, we will assume that the utility derived from a single landing page is equivalent to reading a fraction of each member of a set of articles (where in principle the fraction could be greater than one).

First consider a case without aggregators. For each news type c, the user can utilize direct navigation to outlet home pages, search, and social, and translate 1 unit of time into consumption of a set of articles, denoted π_c^D, π_c^{Se}, and π_c^{So}, respectively. As discussed above, when the user visits outlet home pages, the visit results in consumption of news π_c^H (in units equivalent to articles). Thus, π_c^H describes the quantity of news of each type consumed directly on the home pages, and π_c^D describes the quantity of articles consumed by clicking on links on the home pages. Let $\pi_c = \pi_c^D + \pi_c^{Se} + \pi_c^{So} + \pi_c^H$ denote the sum of these. So far, we have not placed any restrictions on the discovery process; we have simply introduced notation describing its output in terms of the objects that create utility for the consumer. For example, if the consumer selects investments of time in visiting home pages, search engines, and social media to maximize expected utility, our notation can be interpreted as summarizing the number of articles consumed when the consumer follows an optimal time investment policy.

Our next step is to introduce a key simplifying assumption, one that leads to a tractable empirical model as well: we assume that the discovery process has constant returns to scale, so that allocating more or less time results in a proportional increase or decrease in the news discovered of each type from each source. In a more detailed microeconmic model of search, the constant returns assumption would imply restrictions on the process by which users engage in search and discovery, the availability of content and home pages, as well as the technology available to search.

Now consider the case where aggregators are available. We will assume there is a single aggregator, Google News, and that the availability of Google News changes the amount of news a consumer can consume per unit of time, both by changing the number of articles of each type that can be discovered, and by introducing a new home page that includes partial news stories from a large variety of publishers. We denote the consumption of news through direct navigation, search, social, and outlet home pages when Google News is also available as $(\tilde{\pi}_c^D, \tilde{\pi}_c^{Se}, \tilde{\pi}_c^{So}, \tilde{\pi}_c^H)$ with sum $\tilde{\pi}_c$. We denote the news articles accessed through Google News by $\tilde{\pi}_c^{GN}$, and the consumption of news through reading the Google News home page as $\tilde{\pi}_c^{HGN}$. We assume that the aggregator-augmented technology is at least as productive as the general-purpose technology at finding articles: $\tilde{\pi}_c + \tilde{\pi}_c^{GN} + \tilde{\pi}_c^{HGN} > \pi_c$ for all c.

The next definition formally captures the notion of aggregators complementing or substituting for traditional news-reading.

Definition 1 *Aggregators are substitutes (complements) to traditional news-reading technology for news of type c if $\tilde{\pi}_c + \tilde{\pi}_c^{GN} < (\geq)\pi_c$.*

Note, that even though aggregator-augmented technology always weakly increases total news consumption it might decrease the number of news stories that are directly consumed on the publishers' websites if aggregators are substitutes and consumers can effectively consume news directly on aggregators' home pages ($\tilde{\pi}^{HGN}$).

We can refine the definition further by saying that Google News is a substitute (complement) for reading articles if $\tilde{\pi}_c^D + \tilde{\pi}_c^{Se} + \tilde{\pi}_c^{So} + \tilde{\pi}_c^{GN} < (\geq) \pi_c^D + \pi_c^{Se} + \pi_c^{So}$. Analogously, Google News is a substitute for publisher home pages if the consumption of those publisher home pages is lower in a regime with Google News. Note, we have not directly defined notation for the number of page views of publisher home pages, because π^H expresses consumption of home pages in units of articles.

We can now derive consumer demand for news stories of type c with both technologies. Utility maximization implies that when aggregators are not available, a user will consume

$$N_{i,c}^t = \tau^t \tau^i \alpha_{i,c} \pi_c \tag{2}$$

while a user of the aggregator-augmented technology will consume:

$$N_{i,c}^{t*} = \tau^t \tau^i \alpha_{i,c} \left(\tilde{\pi}_c + \tilde{\pi}_c^{GN} + \tilde{\pi}_c^{HGN} \right) \tag{3}$$

Since we will measure news consumption by the number of pages visited and the dwell time spent on publisher's websites we denote the news consumed directly on publishers' websites when using aggregator-augmented technology with $\tilde{N}_{i,c}^t$:

$$\tilde{N}_{i,c}^t = \tau^t \tau^i \alpha_{i,c} \left(\tilde{\pi}_c + \tilde{\pi}_c^{GN} \right) \tag{4}$$

For a particular type of news on a given day, the ratio of news consumed on publisher websites using the two technologies is directly proportional to the relative productivity of the two different discovery processes:

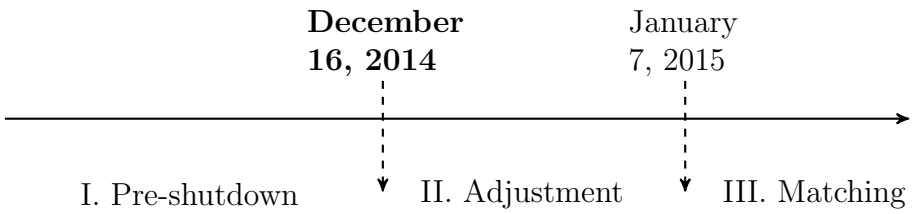
$$\frac{\tilde{N}_{i,c}^t}{N_{i,c}^t} = \frac{\tilde{\pi}_c + \tilde{\pi}_c^{GN}}{\pi_c} \tag{5}$$

2.2 Empirical Implementation and User Matching

We next take this model to our data. We distinguish between three main periods shown in Figure 1 – the pre-shut period I (before December 16, 2014), a 3-week adjustment period (until January 7, 2016) and a matching period (after January 7, 2016). We define a set of "treatment users" as those who have used Google News before the shutdown. For each such treatment user i we find a control user $\hat{i}$ who *does not* use Google News in the pre-shutdown period but consumes news in the same way as the treatment user during the matching period. We will then assume that treatment and control users have the same preferences (e.g. $\tau^i = \tau^{\hat{i}}$ and $\alpha_{i,c} = \alpha_{\hat{i},c}$ for all news types c) because they make the same consumption decisions when having access to the same technology.

We will now describe the details of the matching algorithm.

Figure 1: Timeline for matching Spanish control and treatment users



Active users. Section A-1 of the supplementary appendix describes in detail how we construct our user-level data set from the browser logs. Our dataset consists of a sample of desktop users and we focus on users who are active 90% of the weeks between October 1, 2014 and March 17, 2013 and who read news for at least 10 weeks. The universe of *active users* in Spain comprises about 158,000 people.

Browsing events are linked through referrer URLs: if a page load has a referrer it indicates that the user clicked on a link on that referring page to visit the present page. The browsing stream can be partitioned into a set of trees where the root of each tree either has an empty referrer or refers to a non-publisher page (such as a Google search page, a social media page or a news aggregator). We refer to these trees as "news mini sessions" (or NMS). The root of each NMS defines the referral mode: we distinguish direct navigation (an empty referrer), search (Google, Bing or Yahoo search referral), social (Facebook or Twitter referral) and other referrals.

We identify a "treatment user" as any user who has used Google News at least once during the pre-shutdown period. The remaining users are potential control users.

Topics. Section A-2 of the supplementary appendix describes in detail how we construct topics. We classify all URLs into landing pages and article pages based on frequency distribution of page loads across the observation period (articles tend to have most page loads concentrated within a few days after publication while landing pages are visited at a fairly constant rate throughout).

We then scrape all 110 million article pages and focus on the 61% of articles that contain more than 100 words (the remainder are often slide shows or video pages). We compare the content of each such article to all the Wikipedia articles published on the Spanish-language Wikipedia. We exploit the fact that most major news events receive a Wikipedia entry within days or weeks. Wikipedia provides us with a convenient and stable set of topics. We construct a set of nested topics consisting of a "super-topic" (such as *Health*) and a sub-topic (such as *Spain Ebola crisis* which hit Spain in 2014/2015). We manage to classify 53% of all articles into about 300 topics. The top 50 topics cover 75% of all page loads.

We hired 7 student evaluators to rate the quality of our super-topic and sub-topic assignment for 500 articles and found that 95% of super-topic and 85% of sub-topic assignments were deemed correct.

Table 1: Differences between matched treatment and control users

statistic	is treated?	total	news	article	search	breaking	"hard news"
min	0	173	101	0	0	0	0
min	1	202	64	0	0	0	0
median	0	3,635	293	122	61	33	13
median	1	3,609	292	124	60	32	13
mean	0	5,250.921	396.243	170.920	99.752	50.982	25.011
mean	1	5,339.146	397.523	174.262	103.655	49.960	24.403
sd	0	5,301.106	294.181	144.934	124.428	54.563	36.368
sd	1	6,139.647	296.541	148.091	136.504	52.349	32.624
max	0	61,397	1,368	1,143	1,178	793	462
max	1	140,589	1,372	1,140	1,149	766	295
N	0	2,317	2,317	2,317	2,317	2,317	2,317
N	1	2,317	2,317	2,317	2,317	2,317	2,317

Matching algorithm. For each treatment user we rank all potential control users according to a proximity score that is based on the components listed in Table 1: (1) total overall page views, news page views (landing pages and articles) and news article page views, (2) news page views accessed through search and (3) breaking news page views and page views on hard news (excluding celebrity news and sports among top 50 topics).

Formally, we calculate for each variable the empirical distribution among all active users. Then for each user we calculate its percentile rank in the distribution. For each treatment user we calculate the difference from each potential control user in these percentile scores and weigh these scores equally to calculate a single proximity score. This procedure ranks potential control users for each treatment user. We then use the random serial dictatorship algorithm to assign a unique control user to each treatment user. The results of the matching are displayed in Table 1[6] Overall, the treatment and control group differ by less than 5% along each of the matched dimensions (where means are compared using data only from the matching period).

Empirical model. Our matched samples of treatment and control users allows us to estimate the effect of the Google News shutdown by comparing the news consumption of treatment and control users during the pre-shutdown period. Since we have constructed the control group such that (by revealed preference, under the assumptions of our stylized model) both groups of users have identical preferences, we can interpret the consumption behavior of matched control user as the predicted (counterfactual) behavior for the corresponding treatment user if Google News was not available.

Our approach of matching on behavior *after* the "treatment" (taking away Google News)

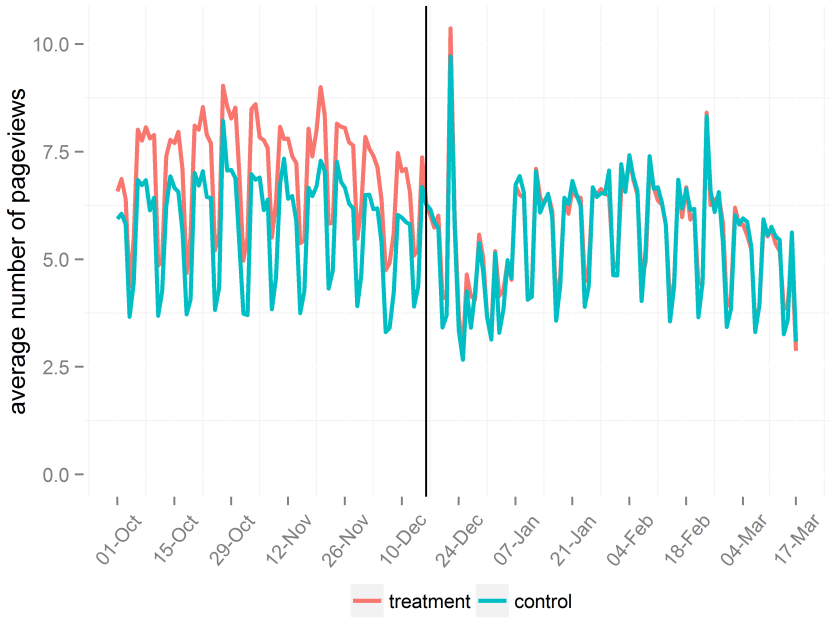
[6]After applying the matching procedure we focus on treatment-control pairs with at least 100 news pageviews in the matching period. However, all our qualitative results hold up even when using the full sample.

has been applied is somewhat non-standard at first glance. A superficially more natural approach might be to match users based on their behavior in the pre-shutdown period. However, on closer examination, it would be hard to justify an empirical strategy based on such an approach. Since in the pre-shutdown period, Google News users are accessing news through a different search and discovery technology, there is no reason to believe that a Google News user and a non-Google News user who have the same preferences would read the same news–in general we would not expect that. On the other hand, two users with the same preferences should read the same amounts of different types of news in the post-shutdown period.

Our analysis would be more straightforward if Google News was introduced as an option rather than taken away. Our approach needs to rely on a few additional assumptions, namely that Google News does not have persistent effects on reader behavior even after it disappears. As described above, to partially deal with the possibility of persistence, we leave out an "adjustment period" after Google News shuts down. However, we do not see evidence in the data that the treatment and control users have differences in behavior that change over time, reducing the potential for concern.

As a robustness check, we also considered an empirical approach based on matching on pre-shutdown news consumption and measuring differences in the post-shutdown period; this approach yielded qualitatively similar results, as we discuss further below.

Figure 2: News consumption of the treatment and control groups over time



3 Google News and Overall News Consumption

We now analyze the impact of Google News on *overall* news consumption across all news categories. The magnitude of this combined effect is fundamental for publishers and policy makers who have to decide whether aggregators steal traffic from news creators (substitutes view) or increase traffic (complements view). In order to gain some intuition, we start with a graphical analysis in Figure 2 which shows the average daily news reading (in term of number of pageviews) over time for both control and treatment users.

As a result of our matching procedure the two groups have very similar news consumption levels after the shutdown (marked with the vertical line). In fact, the consumption levels of both groups are nearly indistinguishable even during the adjustment period. Moreover, control users have an overall stable consumption level before and after the shutdown (other than the holiday period and a spike corresponding to the indictment of Spain's Princess Cristina on corruption charges) which is to be expected since the Google News shutdown did not affect them. The stability also implies that the details of how we model time trends will not have a big impact on our results.

In contrast, Google News users have much higher news consumption compared to control users before the shutdown. Consequently, there is a sharp and pronounced drop in news volume once Google News becomes unavailable. We call this phenomenon the "volume effect."

In order to gain some insight how Google News affects news discovery we show the referral modes for total news reading for both groups in Figure 3. The referral modes are classified as direct navigation, search, Google News and other (which includes social media, emails, forums, etc).

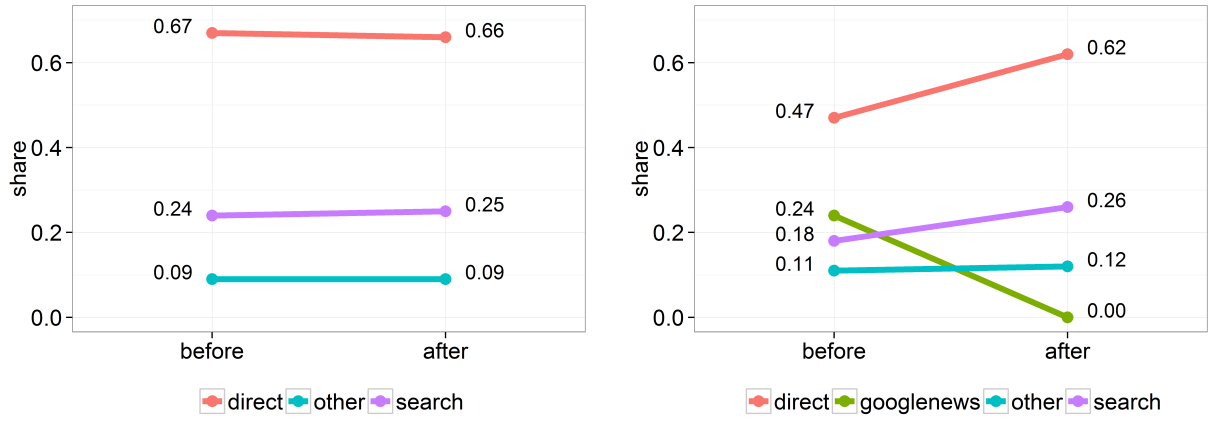
For control users, referral shares are constant in time, as expected. During the post-period, treatment users have very similar referral shares compared to control users. This is again expected since we are matching both on total news consumption and consumption via search. Before the shutdown, reading through Google News makes up 24% of users' consumption. Direct navigation and search take up this share after the shutdown, while the referral share through other modes stays roughly constant.

3.1 Estimating the Volume Effect

We now estimate the effect of removing Google News by comparing the consumption of treatment and control users pre-shutdown. Using our model, we can express the total news consumption of all treatment users during the pre-shutdown period by $\tilde{N}^I$:

$$\tilde{N}^I = \tau^I \sum_{\hat{i},c \in \mathcal{C}} \tau^{\hat{i}} \alpha_{\hat{i},c} \left(\tilde{\pi}_c + \tilde{\pi}_c^{GN} \right) \tag{6}$$

Figure 3: Referral shares, total news consumption.



(a) Control users

(b) Treatment users

The corresponding consumption of all control user is N^I:

$$N^I = \tau^I \sum_{i,c \in \mathcal{C}} \tau^i \alpha_{i,c} \pi_c \tag{7}$$

Therefore, the ratio of total news consumption of treatment and control users equals:

$$\frac{\tilde{N}^I}{N^I} = \frac{\sum_{\hat{i},c \in \mathcal{C}} \tau^{\hat{i}} \alpha_{\hat{i},c} \left(\tilde{\pi}_c + \tilde{\pi}_c^{GN} \right)}{\sum_{\hat{i},c \in \mathcal{C}} \tau^{\hat{i}} \alpha_{\hat{i},c} \pi_c} \tag{8}$$

Note, that we are exploiting our matched sample here which allows us to replace τ^i with $\tau^{\hat{i}}$ and $\alpha_{i,c}$ with $\alpha_{\hat{i},c}$. This ratio measures the overall impact of Google News on treatment users, while factoring out seasonal and day effects.

Taking the logarithm of both sides of (8) suggests an empirical approach to estimating the percentage change in news consumption when Google News is removed. Table 2 shows the results of this calculation, where we average the logarithm of the left-hand side of (8) using data from the pre-period, where we use the number of pageviews as our measure of news consumption. The top row shows the volume effect across all navigation modes for total news as well as the sub-categories of only articles, landing pages and other content (slide shows and videos).[7] Column (5) explicitly excludes Google News landing pages and therefore only captures publishers' landing pages. Rows 2 to 4 capture the change just for direct navigation, search and other modes of accessing news. We calculate standard errors using the bootstrap, sampling treated-control pairs as a unit.

Treatment users have 19.7 percent higher consumption in the pre-shutdown period compared to control users, including their consumption of the Google News home page. As a result of our matching procedure and the exogenous nature of the shutdown we interpret this as

[7]Slideshows and videos are not classified as articles if the word count is less than 100 words.

Table 2: Volume difference measures for news types according to referral shares and pageview types (number of pageviews).

Referral mode	Total	Total (excl. GN)	Article	Landing page	Landing page (excl. GN)	Other content
Total	0.197***	0.096***	0.288***	0.158***	-0.085***	-0.076
	(0.019)	(0.018)	(0.021)	(0.024)	(0.023)	(0.058)
Direct	-0.172***	-0.172***	-0.223***	-0.117***	-0.117***	-0.268***
	(0.025)	(0.025)	(0.030)	(0.026)	(0.026)	(0.077)
Search	-0.056*	-0.073**	-0.023	-0.102***	-0.150***	-0.050
	(0.032)	(0.032)	(0.032)	(0.042)	(0.041)	(0.097)
Other	0.343***	0.340***	0.420***	0.166	0.148	0.158*
	(0.058)	(0.058)	(0.053)	(0.105)	(0.105)	(0.095)

Notes: Bootstrapped standard errors in parenthesis.

Table 3: Volume difference measures for news types according to referral shares and pageview types (dwelltime).

Referral mode	Total	Total (excl. GN)	Article	Landing page	Landing page (excl. GN)	Other content
Total	0.307***	0.170***	0.371***	0.258***	-0.044*	0.105
	(0.022)	(0.019)	(0.024)	(0.032)	(0.026)	(0.071)
Direct	-0.136***	-0.136***	-0.231***	-0.078***	-0.078***	-0.142*
	(0.027)	(0.027)	(0.033)	(0.030)	(0.030)	(0.081)
Search	-0.004	-0.028	0.029	-0.053	-0.118***	-0.018
	(0.036)	(0.035)	(0.040)	(0.047)	(0.046)	(0.124)
Other	0.429***	0.426***	0.460***	0.300***	0.281***	0.424***
	(0.060)	(0.060)	(0.060)	(0.117)	(0.117)	(0.145)

Notes: Bootstrapped standard errors in parenthesis.

the causal effect of the presence of Google News on news consumption volume. This volume change comes from two sources: (1) Google News users consume 28.8 percent more articles but 8.5 percent fewer landing pages (omitting the Google News landing page). Hence, Google News is a complement to overall news reading and articles but a substitute to landing pages. The landing page result is intuitive since Google News directly links to publishers' articles and bypasses the landing pages. However, this decline is more than compensated for by increased traffic to articles. (It should be noted, however, that a large share of advertising revenue comes from the publishers' landing pages.)

Google News has only a small effect on news accessed through search. It complements news accessed through "other" navigation modes (such as social media). However, as we saw before, these make up only a small share of total news referrals.

A question that might arise is whether Google News referrals might be of lower quality than others; perhaps when using Google News, readers will click on many articles, glance at them, and then quickly return to the Google News home page. To assess this hypothesis, Table 3 shows the same calculations using dwelltime (time spent on the page) as a measure of news consumption.[8] The numbers corroborate our previous conclusions using pageviews.

3.2 Volume Effect by Outlet Size

Table 4: Volume effect calculations (number of pageviews).

Outlet type	Referral mode	Total	Article	Landing page	Other content
Top 20	Total	0.022	0.201***	-0.137***	-0.109*
		(0.023)	(0.025)	(0.026)	(0.064)
Top 20	Direct	-0.218***	-0.261***	-0.169***	-0.282***
		(0.031)	(0.037)	(0.030)	(0.085)
Top 20	Search	-0.120***	-0.076**	-0.193***	-0.045
		(0.038)	(0.036)	(0.048)	(0.092)
Below top 20	Total	0.263***	0.446***	0.046	0.106
		(0.037)	(0.035)	(0.052)	(0.100)
Below top 20	Direct	-0.042	-0.113**	0.012	-0.111
		(0.051)	(0.054)	(0.055)	(0.124)
Below top 20	Search	0.028	0.061	-0.021	-0.044
		(0.059)	(0.056)	(0.082)	(0.132)

Notes: Bootstrapped standard errors in parenthesis.

We now break out the volume effect by outlet size. This is an important exercise because the Spanish copyright reform was primarily advocated by bigger newspapers, and also because theory suggests that the effect might be different, with the search and discovery function

[8]Section A-1.2 of the Data Appendix explains how the dwelltime variable is logged in our data.

playing a more important role for smaller outlets. For smaller outlets, the users may not expect the benefit of accessing the landing pages to outweigh the time to access it, and further, the users may not be aware of some smaller publishers. Using the whole sample period and all active users, we order news outlets based on their size measured in terms of total pageviews. Then we compare the top 20 publishers to the smaller ones. Results are shown in Table 4.

A striking pattern can be observed when we compare the total news consumption of treatment and control users: while the effect of Google News on the top 20 publishers is not statistically significantly different from zero (with estimates small in magnitude), smaller outlets gain as much as 26.3 percent from the presence of Google News. When we dissect this effect by page view type, we see that the availability of Google News changes the page view mix of towards articles and away from landing pages for larger outlets, but the two effects cancel in the aggregate. For smaller outlets, the landing page traffic is unaffected by Google News but article pageviews increase by 44.6 percent. Consistent with the promotion of smaller outlets, Google News also makes it easier for users to facilitate multiple news sources: Table 5 compares the user-level Herfindahl indices for outlets for the top 20 topics by referral mode. For each topic, the Google News concentration index is lower than for direct navigation.

Table 5: Mean of user-level Herfindahl indices for outlets within a topic (pre-period, treatment users).

Topic	direct	googlenews	search
Business/Finance: Macroeconomics and Economic Policy	0.793	0.603	0.789
Politics: Independence of Catalonia	0.797	0.531	0.793
Spain: Government	0.770	0.562	0.800
Health: Spain Ebola Crisis	0.825	0.558	0.785
Sports: Atlético de Madrid	0.834	0.721	0.863
Entertainment: TV Show Gran Hermano VIP	0.900	0.741	0.806
Spain: Corruption Operation Púnica	0.839	0.676	0.856
Celebrities: Duchess of Alba dies	0.874	0.708	0.808
Sports: FC Barcelona	0.865	0.720	0.891
Sports: Real Madrid	0.884	0.705	0.873
Sports: Soccer Players	0.886	0.726	0.876
International News: Central and South America	0.855	0.712	0.898
Sports: Formula 1	0.882	0.734	0.822
Entertainment: Movies and Actors	0.900	0.811	0.889
Entertainment: Television	0.893	0.784	0.881

Therefore, the evidence supports the hypothesis that Google News is neither a complement or substitute for bigger publishers, but is a substitute for the high-revenue landing pages of those outlets. We also see that Google News is a substitute for articles accessed through direct navigation–highlighting that large outlets lose some of their curation role. If part

of the long-term incentive for news outlets to maintain their brand comes from the way they curate news, through the selection of articles to highlight prominently on their landing pages, then the fact that Google News is in effect selecting what articles from each outlet to highlight on the Google News home page may decrease the incentive of publishers to invest in the quality of their curation and thus their brand. This is an example of a broader concern publishers articulate surrounding aggregators and social media: they worry that they are being "disintermediated" and "commoditized,"[9] consistent with a decreased ability to differentiate their products in the eyes of consumers, as their content is accessed instead through an intermediary.

On the other hand, our evidence supports the hypothesis that Google News is a strong complement for small outlets. This implies that bigger outlets gained compared to smaller ones due to the shutdown of Google News and users visited a less diverse set of outlets. The social welfare implications of this finding are unclear; there are potentially competing effects. If small outlets produce unique content with alternative viewpoints or reporting, then Google News supports "media diversity" and helps smaller publishers get viewership and thus returns on their investments in journalism. On the other hand, if smaller outlets mostly copy news from larger outlets, without investing in reporting, then Google News might decrease the returns to investment for the primary sources of investigative journalism. This paper does not provide evidence that speaks directly to these welfare tradeoffs. In principle, the investments in journalism can be measured using data from newspaper employment and expenditures, and diversity of viewpoints as well as the extent of original reporting can be assessed from textual analysis. This represents an interesting avenue for future work.

Table 6 repeats the same exercise using dwelltime as the consumption measure. Overall, the evidence supports the findings from our analysis of pageviews. The main difference is that total dwelltime increases slightly even for bigger outlets under the presence of Google News: this reflects the fact that users tend to spend more time on article pages compared to landing pages. Dwell time on landing pages declines for large outlets, in line with the results for page views. In the current advertising environment, advertising is typically sold on the basis of page views, not dwell time, so the newspapers may not see revenue increases that correspond to the dwell time increase.

4 Decomposing the Volume Effect

In Section 3 we analyze the overall drop in news consumption when Google News becomes unavailable. In order to better understand *why* consumption decreases we examine how reading changes as a function of news characteristics. Based on anecdotal evidence, we focus on five characteristics: recency of news/breaking news, the extent to which different news stories are widely covered by newspapers (global supply scarcity), the extent to which news stories are well covered by readers' favorite news outlets (relative supply scarcity), popularity of news stories, and whether a story is about "hard" or "soft" news. As motivated in the

[9]See, e.g., Sterling (April 7, 2009; accessed November 17, 2016), Filloux (July 27, 2016; accessed November 17, 2016), Weiner and Group (May 11, 2015; accessed November 17, 2016)

Table 6: Volume drop calculations (dwelltime).

Outlet type	Referral mode	Total	Article	Landing page	Other content
Top 20	Total	0.103***	0.308***	-0.106***	0.100
		(0.024)	(0.029)	(0.031)	(0.081)
Top 20	Direct	-0.189***	-0.279***	-0.137***	-0.157*
		(0.032)	(0.040)	(0.036)	(0.089)
Top 20	Search	-0.086**	-0.019	-0.195***	0.004
		(0.042)	(0.046)	(0.054)	(0.135)
Below top 20	Total	0.324***	0.489***	0.123***	0.132
		(0.038)	(0.040)	(0.053)	(0.124)
Below top 20	Direct	0.003	-0.111*	0.075	-0.053
		(0.053)	(0.065)	(0.056)	(0.178)
Below top 20	Search	0.108	0.113*	0.121	-0.214
		(0.069)	(0.067)	(0.104)	(0.181)

Notes: Bootstrapped standard errors in parenthesis.

introduction, understanding what kinds of news are particularly affected by Google News helps explain how Google News competes with or complements existing outlets, and also sheds light on how aggregators affect the incentives of news publishers to produce different types of news.

Formally, we define each news category by whether the category does or does not have each of the D "characteristics" (in our empirical work, the $D = 5$ characteristics just described), so that c is a vector in $\{0,1\}^D$. For each of the $d = 1, .., D$ conditions, there is a corresponding component of c, denoted $c_d \in \{0,1\}$, that indicates whether news type c satisfies condition d. For example, d might represent the condition "the news is breaking news" or "the news is hard news," and $c_d = 1$ if category c satisfies the condition.

4.1 News Characteristics

We now define the characteristics we use in our empirical analysis.

4.1.1 Breaking News

Google News is an algorithmic service that is capable of collecting the most recent stories from multiple news sources. In fact, as of 2016 there is a "See realtime coverage" button on Google News that provides users with the latest relevant news for a particular event. One might therefore suspect that aggregator-augmented browsing lowers the cost for consuming breaking news compared to traditional news browsing. Most traditional readers rely only on a handful of newspapers that (1) may happen to miss the latest events and (2) do not provide links to other news sources covering the same events. As a result, we hypothesize that the

shutdown of Google News leads to a disproportionate drop in the consumption of *breaking news* stories. We proxy breaking news by determining the hour of publication of each news article and measuring the time elapsing between the publication and the consumption of the news story.

4.1.2 Supply Scarcity

Some news topics are covered by a large number newspapers while others receive less attention. This mismatch can be caused by newspapers miscalculating readers' interest or by a lack of reporters who can write such stories. For example, a terrorist attack in a neighboring country might be of great interest to local readers, but newspapers might lack correspondents in that country. We distinguish between two types of scarcity: *relative supply scarcity* captures the idea that users read a limited number of publications and these publications might poorly cover certain news topics compared to other newspapers. In contrast, *global supply scarcity* captures poor coverage of a certain topic across all newspapers.

Google News can potentially mitigate both types of scarcity. First of all, personalization allows aggregators to adapt to user preferences. For example, if a user's favorite newspapers poorly cover financial news then Google News can adapt accordingly and prioritize financial news. Moreover, Google News provides typically multiple references for each news topic which makes it easier for users to find related coverage for globally scarce news topics.

We next formally define both types of scarcity.

Relative Supply Scarcity. Let us index users by i, topics by d, newspapers by n and days by t. The daily share of articles in newspaper n on topic d on day t is denoted with s_{ndt}.[10] The total pageviews consumed by user i for newspaper n during the matching period is denoted with y_{in} – it measures the intensity with which user i reads newspaper n. We define $y_i = \sum_n y_{in}$ as the total pageviews of user i across all newspapers during the matching period and $y_n = \sum_i y_{in}$ as the total pageviews for newspaper n across all users during the matching period.[11]

We then define *aggregate supply* for topic d on day t:

$$x_{dt} = \frac{\sum_n y_n s_{ndt}}{\sum_n y_n} \tag{9}$$

Intuitively, aggregate supply measures the average coverage for news topic d across all newspapers weighted by their popularity.

We then define the *user-level supply* of news faced by user i for topic d on day t:

$$x_{idt} = \frac{\sum_n y_{in} s_{ndt}}{\sum_n y_{in}} \tag{10}$$

[10]Note, that $\sum_d s_{ndt} = 1$.

[11]Note: we focus on direct page-views, all pages (both landing and article pages).

Intuitively, take a random pageview for the user - with probability y_{in}/y_i the user selects newspaper n and then reads a random article which is going to be on topic d with probability s_{ndt}. This is the reader's individual-level supply of articles on topic d on day t.

This allows to then define *relative supply* r_{idt} as follows:

$$r_{idt} = x_{idt} - x_{dt} \tag{11}$$

Relative supply has two important aggregation properties. First of all, for any specific user, relative supply across topics sum up to 0:

$$\sum_d r_{idt} = 0 \tag{12}$$

Second, for any specific topic, relative supply across users sum up to 0:[12]

$$\sum_i w_i r_{idt} = 0 \quad \text{where } w_i = \tfrac{y_i}{\sum_i y_i} \tag{14}$$

Therefore, topics are neither scarce nor abundant *on average* but they are only so *within* users.

Global Supply Scarcity. We identify globally scarce topics by comparing aggregate demand to aggregate supply x_{dt}. Formally, we define aggregate demand q_{dt} as the pageview share across all topics:

$$q_{dt} = \frac{y_{dt}}{\sum_d y_{dt}}$$

Relatively globally scarce topics are defined as those topics d for which

$$q_{dt} - x_{dt} > 0$$

Intuitively, these topics have excessively high readership compared to the number of articles (information) that are available through traditional news browsing (direct navigation).

4.1.3 Popularity and Hard News

Google News might also promote popular news topics. We say that a topic is popular if it among the top 5 topics measured by pageviews during the pre-shutdown period.

Google News might be particularly effective in lowering the cost of finding hard news and niche topics which are insufficiently covered by most news outlets. Therefore, we might expect a larger volume effect for hard news topic when Google News is no longer available. Formally, we say that a topic covers hard news if it is neither a celebrity or sports topic.

[12] First of all, we know that $\sum_i w_i = 1$. Therefore, we have $\sum_i w_i x_{dt} = x_{dt}$. Moreover, we have:

$$\begin{aligned} \sum_i w_i x_{idt} &= \sum_i \frac{y_i}{\sum_i y_i} \frac{\sum_n y_{in} s_{ndt}}{\sum_n y_{in}} = \sum_i \frac{\sum_n y_{in} s_{ndt}}{\sum_i y_i} \\ &= \sum_n \frac{\sum_i y_{in} s_{ndt}}{\sum_n y_n} = \sum_n \frac{y_n s_{ndt}}{\sum_n y_n} = x_{dt} \end{aligned} \tag{13}$$

4.2 Differences Between Treatment and Control Users

Our matching algorithm does not explicitly match treatment and control users' consumption along all characteristics; the only two incorporated in the matching are breaking news and hard news. Thus, testing whether news consumption differs in other characteristics helps evaluate whether there are important residual sources of heterogeneity between the two groups. However, Table 7 demonstrates that treatment and control users look very similar along the five basic characteristics in the matching period; in the Appendix, we test equality all $2^5 = 32$ categories defined by the vector of characteristics c. To account for multiple testing, we report whether the differences are significant using the Benjamini-Hochberg (BH) procedure (where we apply the procedure to the set of 5 characteristics, and then separately to the set of 32 categories, in each case reporting whether we can reject equality of the two groups at the 10% level).

Results for the 5 characteristics are reported in Table 7, while results for the 32 categories are in Appendix Table 10. Among the largest differences we find, we see that globally scarce news consumption is 6.6% higher for control users, and the difference is statistically significant (when the hypothesis test is considered in isolation) at the 5% level. We also see differences two of the 32 categories, namely categories $0:1:0:0:1$ and $0:0:0:0:1$ (globally scarce and breaking, and only breaking, respectively), where we reject equality of the news browsing volumes even after applying the (BH) correction. Although breaking news was included as a feature for matching, other characteristics were also considered and so we did not attain perfect matching in practice. The category that only has the globally scarce characteristic also shows fairly large discrepancies between treated and control, although the test for equality is not significant after applying the BH correction. The other categories are similar in magnitude and differences are not significant.

Table 7: Differences between treatment and control users on news types (top 50 topics). To deal with issues of multiple testing, the Benjamini-Hochberg procedure was used to control for the false discovery rate at the 10% level.

characteristic	treatment	control	p-value	BH corrected
Relative scarcity	15.691	15.506	0.694	not rejected
Global scarcity	24.625	26.258	0.031	not rejected
Popularity	15.387	16.255	0.129	not rejected
Hard news	24.275	24.897	0.482	not rejected
Breaking news	12.294	13.119	0.062	not rejected

4.3 Referrals from Google News versus other Referral Modes, by Characteristic

Next, we show how news consumption obtained through Google News differs from other referrals modes along our five characteristics.

Figure 4: Referral shares and news reading volume as a function of time after publication

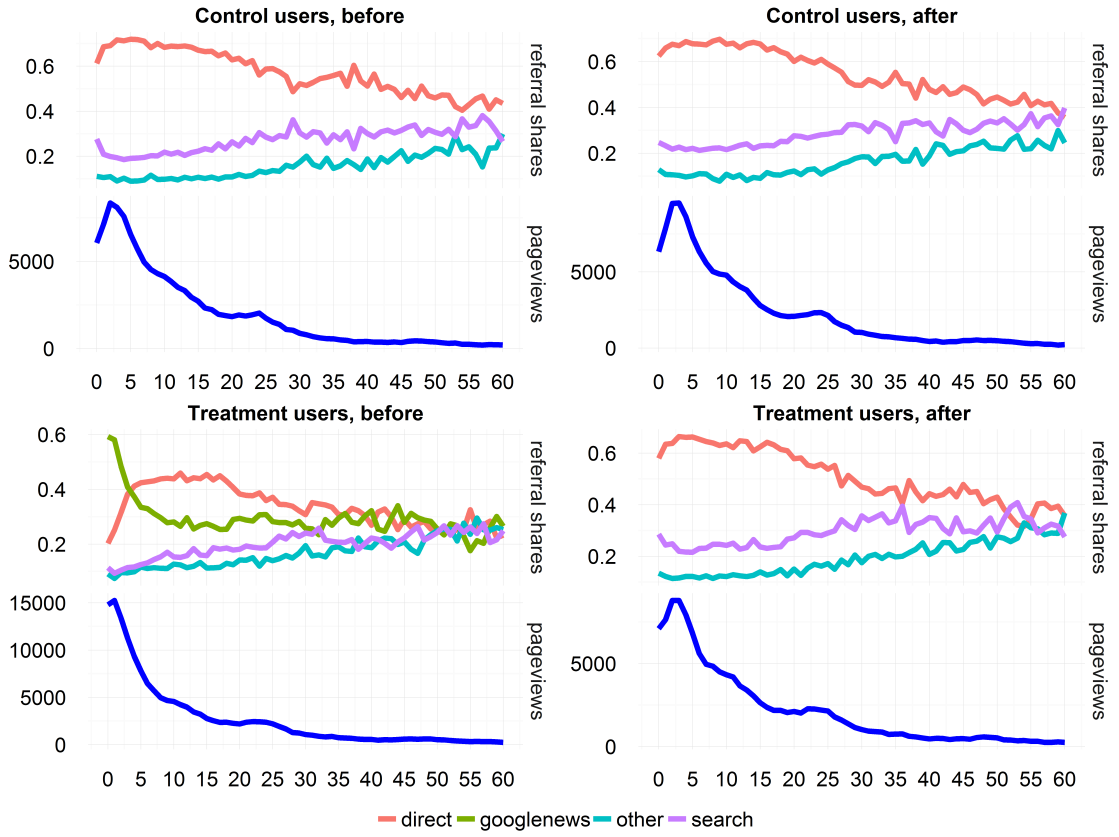


Table 8: Share of page views originating from different referral modes by User-scarcity for treatment users pre-shutdown (top 50 topics)

referral mode	non-scarce	scarce
direct	0.397	0.222
googlenews	0.315	0.436
other	0.130	0.158
search	0.157	0.183

First, Figure 4 shows the number of pageviews and share of news consumed through the four main referral modes for treatment and control users. As expected, control users' news consumption patterns are similar before and after the shutdown. For control users (as well as post-shutdown, for treatment users), most news articles are browsed at 3 hours after publication of the articles. Treatment users, however, access news faster after their publication when Google News is available (pre-shutdown): the peak in news consumption occurs at 0 and 1 hours after publication. The Figure also illustrates that Google News is the main referral source for very recently breaking news in the pre-shutdown period. After the shutdown, the peak is very close to that of control users, and it is occurring later than before the shutdown, illustrating the Google News users do not find an alternative source for late-breaking news.

Next, Table 8 presents evidence that Google News is an important source of user-scarce news: its share is much higher for scarce than for non-scarce topics (44% versus 32%). We also verified that this result holds for both light and heavy Google News users, using various cutoffs to define light and heavy.

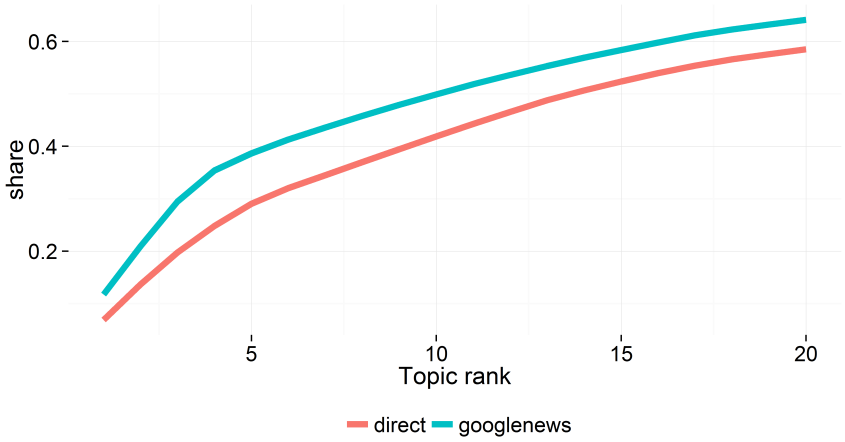
Treatment users also read more popular topics through Google News compared to direct navigation. Figure 5 shows the cumulative reading shares for top topics by referral mode (see Tables 11a and 11b in the Appendix for a detailed breakdown).

4.4 Volume Effect By News Characteristics

We now compare treatment and control users in the pre-shutdown period. Recall, that we matched for each Google News user a corresponding control user who has the same news consumption preferences (through matching in the matching period when both groups have access to the same news browsing technology). Therefore, we can simply compare the consumption of treatment and control users in the pre-shutdown period to simulate the effect of a removal of Google News.

In Figure 6 we depict the change in news consumption between treatment and control users by hour of publication (capturing the breaking news effect) and by scarce and non-scarce topics (red versus blue). The lightly shaded bars (light red and light blue) indicate the decrease in news consumption that would occur if treatment users would simply stop using

Figure 5: Cumulative reading shares of top topics for referral modes direct navigation and Google News.



Google News and not substitute to other publishers; this is calculated by simply removing all news viewing that is referred by Google News from the treatment users' aggregate viewing. We refer to this as the "no-substitution" counterfactual. The darkly shaded decreases (dark red and dark blue) indicate the actual decreases.

First, we observe that the overall (actual) decrease in news consumption that we observed in Section 3 is reflected in every single combination of hour of publication and scarcity. However, the decrease is particularly stark for breaking news and scarce news: scarce breaking news falls by almost 70% while non-scarce, non-breaking news falls by only less than 20%.

Second, comparing the actual decrease to the no-substitution counterfactual, we find that there is very little user substitution for user-scarce news in the first 2 hours of publication: the overall decrease in news consumptions closely tracks the Google News predicted decrease. This finding highlights an important source of differentiation for Google News, as we see that the users are not able to find a good substitute for this news after the shutdown. In general, the figure illustrates that substitution is stronger for non-scarce news. This is consistent with users switching from Google News to direct navigation, since it is easier to find non-scarce than scarce news in a direct way.

Similarly, Figure 7 compares the change in news consumption for popular (blue) and hard news (red). The contrast between popular and unpopular topics for the actual volume drop is not large in magnitude, while the decreases for hard news are larger than those for celebrity and sports news. Except for popular, celebrity and sports news, where there appears to be little substitution, there is moderate substitution for Google News in the other categories (but by no means full substitution, which would lead to decreases of zero).

All of the differences between counterfactual and actual news decreases in the two figures are statistically significant at the 5% level (even after applying a BH correction for multiple testing), except for the differences corresponding to non-scarce news read in 4-10 hours after publication.

Figure 6: No-substitution counterfactual decrease and actual volume decrease by breaking news and user scarcity.

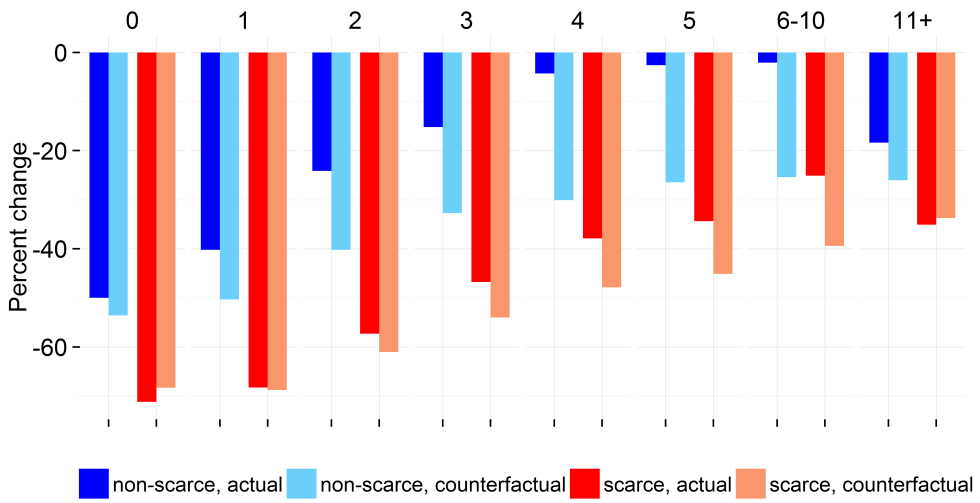


Figure 7: No substitution counterfactual decrease and actual volume decrease by popularity and broad topic.

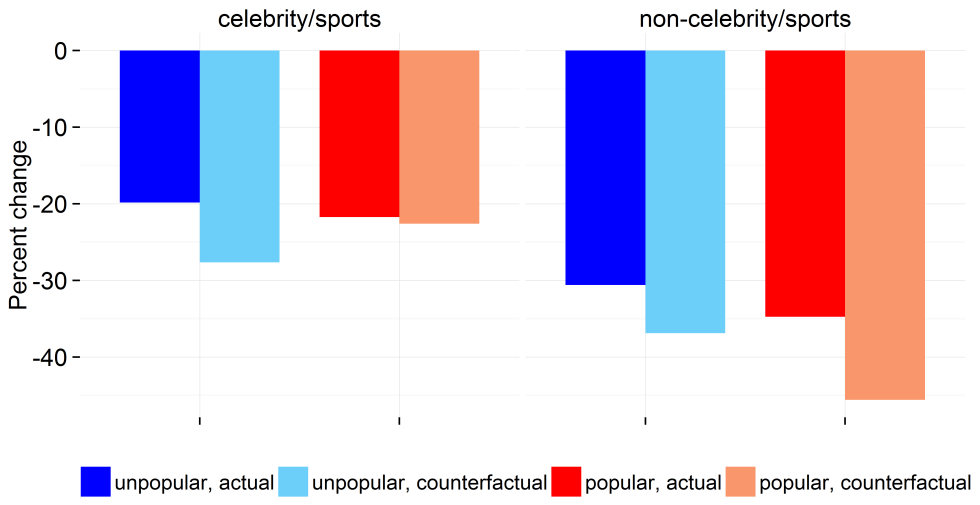


Table 9: Estimating the log-linear structural model for volume drop by news characteristic

	Dependent variable:
	log(control / treatment)
Relative scarcity	0.302^{***}
	(0.035)
Global scarcity	0.013
	(0.020)
Popularity	0.057^{**}
	(0.026)
Hard news	0.138^{***}
	(0.038)
Breaking news	0.222^{***}
	(0.022)
constant	0.005
	(0.041)
Note:	*p<0.1; **p<0.05; ***p<0.01

4.5 Decomposition

In this section we decompose the volume drop of news types (recalling that a category is a vector of binary indicators for news characteristics) into the impact of each of the underlying characteristics. Our empirical strategy is based on comparing the relative consumption of treatment and control users within a news category:

$$\frac{\tilde{N}^I_{i,c}}{N^I_{\hat{i},c}} = \frac{\tilde{\pi}_c + \tilde{\pi}^{GN}_c}{\pi_c} \tag{15}$$

The right-hand side of this equation measures the relative productivity of traditional and aggregator-augmented news browsing when reading news of type c.

In order to decompose the right-hand side of (15), we assume the following functional form:

$$\frac{\tilde{\pi}_c + \tilde{\pi}^{GN}_c}{\pi_c} = \exp(\gamma_0 + \sum_d \gamma_d \cdot c_d + \epsilon_c) \tag{16}$$

Hence, γ_d captures the efficiency loss from losing aggregator access for news dimension d. The constant γ_0 captures the efficiency loss that is not captured by any of the other characteristics.

We estimate the relative contribution of each news characteristic based on the structural model given in (16):

$$\log\left(\frac{N^I_{i,c}}{\tilde{N}^I_{\hat{i},c}}\right) = \gamma_0 + \sum_d \gamma_d \cdot c_d + \epsilon_c \tag{17}$$

Since we have defined $D = 5$ binary news characteristics, there are $2^5 = 32$ news types. We can use the volume drops for those 32 news types to estimate these 6 coefficients.

The estimation procedure we use is a minimum distance procedure. For each news type, we use the sample analogue of equation 17 as a moment condition; details of the method are described in Cameron and Trivedi (2005, p. 202).

We report the results in Table 9. As observed in the previous Section, breaking news and user scarcity are the most important factors in explaining the volume drop – they decrease page consumption by 22.2% and 30.2%, respectively. Hard news and popularity also contribute (13.8% and 5.7%), while global scarcity has a small coefficient which is not significantly different from zero. Interestingly, the constant term is small and not statistically significant which implies that the five characteristics that we have included in our simple additive structural model can fully account for the volume drop.

5 Conclusion

A general theme of innovation since the advent of the internet, starting with eBay and extending to firms like Airbnb and Uber, is that digital intermediaries can reduce search

costs and increase the ability of small, flexible sellers to access consumers. The consumer benefits enabled by these intermediaries are clearly very large in magnitude: consumers can access a much larger diversity of products, potentially at lower prices, and small sellers can be found by consumers, potentially creating large welfare gains when consumers find the perfect match for them. These benefits have been the subject of a literature on the welfare benefits of giving consumers access to the "long tail" of products.

Incumbents and regulators have noted a number of challenges when considering regulation of digital intermediaries. The new entrants often have different business models than incumbents, and may appear in some respects to be in the same product market as incumbents (substitutes), while in other ways they appear to be upstream (complements). For example, eBay can compete with traditional retailers, but it can also help small retailers acquire additional customers or sell unsold inventory; Uber competes directly with taxis and limosines, but can also refer customers to existing limosine businesses who have spare capacity; and AirBnb competes with hotels but can help small lodging providers find consumers. In many cases, the new entrants require less fixed cost investment and expenditures in developing the end consumer product, instead investing research and development to provide sophisticated search and discovery technology. Of course, the value of search and discovery is capped at the value of the products that are available to discover, and so the ultimate welfare evaluation of intermediaries must take into account the long-term quality, assortment, and pricing of final products for consumers. However, economists have long understood that changes in short-term competitive conditions have ambiguous effects on long-term investments in research and development, product quality, and industry structure. Thus, every industry requires its own evaluation.

In the case of News Aggregators, the news industry has called for regulation, resulting in a number of policy interventions across a variety of countries, including the regulatory action in Spain that was the focus of this paper. Most recently, in September of 2016, the European Union has proposed new regulations for news aggregators, as they consider requiring internet companies to pay for news.[13] The empirical evidence we have presented in this paper speaks to some, but not all, of the issues at stake.

Our analysis documents a large, positive effect of Google News on small outlets, as well as on the ability of consumers to access certain types of news, such as breaking news or news that is not well covered on their favorite outlets. These findings highlight the large potential for welfare benefits from improved search and discovery, the "upstream" or complementary role for an intermediary. At the same time, our findings also highlight that while large publishers may not see an effect in overall page views as a result of aggregators, they may lose traffic to their home pages, as well as their role in curating news, as readers read articles referred by Google News at the expense of articles referred by their own home pages (where newspapers monetize the home pages much better than articles). If readers do not pay attention to the identity of the publisher when they read articles on Google News, then the large publishers may lose their incentives to maintain a reputation for quality, and consumers may be less willing to subscribe to the publisher or use the publisher's mobile application.

[13] See, e.g., Schechner and Woo (September 14, 2016; accessed November 17, 2016)

Further research is required to assess the ultimate welfare costs and benefits. More broadly, our analysis covered only a few weeks before and after the change; ideally, we would want to understand the long-term response of both readers and publishers to changes in policy surrounding aggregators.

References

Barthel, Michale, *Newspapers: Fact Sheet*, Pew Resarch Center, June 15, 2016; accessed November 17, 2016. http://www.journalism.org/2016/06/15/newspapers-fact-sheet/.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, 2008, *2008* (10), P10008.

Calzada, Joan and Ricard Gil, "What do News Aggregators Do? Evidence from Google News in Spain and Germany," September 2016. SSRN working paper, available at https://ssrn.com/abstract=2837553.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometric: Methods and Applications*, Cambridge University Press, May 2005.

Chiou, Lesley and Catherine Tucker, "Content aggregation by platforms: The case of the news media," 2015. NBER working paper 21404.

Filloux, Frederic, *News Publishers' Facebook Problem*, Monday Note, July 27, 2016; accessed November 17, 2016. https://mondaynote.com/news-publishers-facebook-problem-6752f1c35037.

Jafri, Salma, *How to Write Headlines Google Will Love & You and I Will Click, Read, and Share*, Search Engine Watch, January 27, 2014; accessed November 17, 2016. https://searchenginewatch.com/sew/how-to/2325076/how-to-write-headlines-google-will-love-you-and-i-will-click-read-and-share.

Kleinberg, Jon and Steve Lawrence, "The structure of the Web," *Science*, 2001, *294* (5548), 1849–1850.

Kleinberg, Jon M, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, 1999, *46* (5), 604–632.

Lu, Kristine and Jesse Holcomb, *Digital News Revenue: Fact Sheet*, Pew Resarch Center, June 15, 2016; accessed November 17, 2016. http://www.journalism.org/2016/06/15/digital-news-revenue-fact-sheet/.

Schechner, Sam and Stu Woo, *EU Executive Proposes New Copyright, Communications Laws*, Wall Street Journal, September 14, 2016; accessed November 17, 2016. http://www.wsj.com/articles/eu-leaders-propose-new-copyright-communications-laws-1473850691.

Smith, Shawn, *Headline writing: How to write web headlines that catch search engine spiders*, New Media Bytes, April 4, 2008; accessed November 17, 2016). http://www.newmediabytes.com/2008/04/10/how-to-write-headlines-for-search-engines/.

Sterling, Greg, *Amid Tensions Google's Eric Schmidt Addresses Newspaper Conference*, Search Engine Land, April 7, 2009; accessed November 17, 2016. http://searchengineland.com/amid-tensions-googles-eric-schmidt-addresses-newspaper-conference-17237.

Weber, Matthew S and Peter Monge, "The flow of digital news in a network of sources, authorities, and hubs," *Journal of Communication*, 2011, *61* (6), 1062–1081.

Weiner, Corey and Jun Group, *Facebook's new "Instant Articles" program poses dilemma for publishers*, VentureBeat, May 11, 2015; accessed November 17, 2016. http://venturebeat.com/2015/05/11/facebooks-new-instant-articles-program-poses-dilemma-for-publishers/.

Appendix

Table 10: Differences between treatment and control users on news types (top 50 topics), $2^5 = 32$ news categories. The types are 1 if a property applies to it and 0 if not, in the order of: relative scarcity, global scarcity, popularity, hard news, breaking news. We report whether we can reject equality of the two groups at the 10% level after applying the Benjamini-Hochberg correction.

value	treatment	control	p-value	BH corrected
0:1:0:0:1	1.272	1.513	0.003	rejected
0:0:0:0:1	1.125	1.314	0.005	rejected
0:1:0:0:0	2.909	3.307	0.011	not rejected
0:1:1:1:0	3.141	3.495	0.068	not rejected
0:1:1:0:1	0.393	0.482	0.084	not rejected
0:1:1:1:1	1.483	1.661	0.111	not rejected
1:0:1:0:0	0.680	0.578	0.119	not rejected
1:0:1:1:1	0.293	0.269	0.132	not rejected
0:0:1:1:1	0.666	0.726	0.144	not rejected
1:0:0:1:0	1.768	1.660	0.154	not rejected
0:0:1:0:1	0.689	0.603	0.189	not rejected
0:0:0:1:0	2.177	2.076	0.253	not rejected
0:1:1:0:0	1.038	1.204	0.257	not rejected
1:0:0:0:1	0.441	0.469	0.260	not rejected
1:0:0:0:0	1.464	1.391	0.282	not rejected
1:1:1:0:1	0.213	0.240	0.325	not rejected
0:0:1:1:0	1.404	1.479	0.341	not rejected
1:1:1:1:0	1.649	1.716	0.369	not rejected
1:0:1:0:1	0.273	0.255	0.382	not rejected
0:1:0:1:1	1.695	1.766	0.402	not rejected
1:0:0:1:1	0.547	0.567	0.484	not rejected
1:1:0:1:1	0.725	0.748	0.494	not rejected
1:1:1:0:0	0.823	0.879	0.499	not rejected
1:1:0:1:0	2.791	2.737	0.603	not rejected
1:1:0:0:1	0.516	0.530	0.612	not rejected
1:1:0:0:0	2.133	2.092	0.634	not rejected
0:1:0:1:0	3.223	3.265	0.771	not rejected
0:0:0:1:1	1.340	1.354	0.801	not rejected
0:0:1:0:0	1.268	1.293	0.827	not rejected
0:0:0:0:0	2.198	2.209	0.933	not rejected
1:0:1:1:0	0.752	0.753	0.976	not rejected
1:1:1:1:1	0.624	0.623	0.991	not rejected

Table 11: Share of pageviews in top topics by referral mode, treatment users.

(a) direct navigation

Topic	Share	Cum. share
Politics: Independence of Catalonia	0.069	0.069
Business/Finance: Macroeconomics and Economic Policy	0.068	0.137
Spain: Government	0.061	0.197
Sports: Atlético de Madrid	0.051	0.248
Health: Spain Ebola Crisis	0.042	0.290
Spain: Corruption Operation Púnica	0.030	0.320
Celebrities: Duchess of Alba dies	0.025	0.345
International News: Central and South America	0.025	0.370
Sports: Real Madrid	0.025	0.394
Sports: FC Barcelona	0.025	0.419
Sports: Soccer Players	0.024	0.443
Sports: Formula 1	0.023	0.466
Entertainment: TV Show Gran Hermano VIP	0.022	0.488
Spain: ETA	0.019	0.507
Business/Finance: Spain Banks	0.017	0.524

(b) Google News

Topic	Share	Cum. share
Politics: Independence of Catalonia	0.117	0.117
Spain: Government	0.093	0.210
Health: Spain Ebola Crisis	0.085	0.295
Business/Finance: Macroeconomics and Economic Policy	0.060	0.354
Spain: Corruption Operation Púnica	0.033	0.387
Business/Finance: Spain Banks	0.026	0.413
Celebrities: Duchess of Alba dies	0.023	0.436
Sports: Atlético de Madrid	0.022	0.458
Sports: Real Madrid	0.021	0.479
International News: Central and South America	0.020	0.499
Sports: FC Barcelona	0.019	0.518
Conflict: War against Islamic State	0.018	0.536
Sports: Soccer Players	0.017	0.553
Spain: ETA	0.016	0.569
Politics: Spain	0.015	0.584

Data Appendix: "The Impact of Aggregators on Internet News Consumption "

Susan Athey
Stanford Business School and NBER

Markus Mobius
Microsoft Research, University of Michigan and NBER

Jeno Pal
Central European University

November 2016

A-1 Browsing Data Processing

In this Section, we describe how we construct a clean dataset from the raw browser log.

A-1.1 Active Users

We only observe desktop users. We restrict attention to users whom we consistently observe over our observation period. User attrition occurs for various reasons such as change in default browser or a browser update that deletes the anonymous machine ID that links all browsing events for the same user across time. We label a user as *active* if both of the following criteria are met: (a) the user has some browsing activity in at least 90 percent of all weeks between October 1, 2014 and March 17, 2015; and (b) there is news reading activity in any 10-week long sub-period (for any of the top 177 Spanish publications).[1] There are 158,575 users who satisfy this definition of activity and they constitute the user base that we use for analysis.

A-1.2 Browsing Stream

For each active user, we observe a set of *browsing sessions* which have a discrete beginning and end time (such as logging into and out of the computer). When there is a gap of more

[1]We constructed this list by looking at all the unique domains that receive referrals from Google News Spain up to December 16, 2014. We then manually double-checked that this list contains the major Spanish newspapers by circulation, top weekly magazines and top radio stations and news portals.

than one hour between browsing events, we define a new session starting after the gap ends. Within a session, we observe a time-stamped sequence of page loads which are linked through a referrer URL. For example, if a user navigates first to the NYT homepage and then clicks on an article on the home page, the browsing stream will show two page loads and the referral source for the second load will be the index page.

An absent referrer field indicates that the user either used a bookmark or typed the URL directly into the browser. An important special case are search queries that result in a visit to a landing page (such as `http://www.nytimes.com`). Often, the user first searches for a phrase such as "NYT" and then clicks on the link that appears on the search page. Even though the referrer field in this case indicates that the page load originated from a search we treat the referrer field as empty (as if the user had clicked a bookmark).

The browsing stream also provides us with a measure of how long the user spent on a certain page ("dwell time"). This data is capped at 300 seconds since larger values typically indicate inactivity (the user has walked away from the computer but left the browser open).

A-1.3 Canonical URLs

URLs are often not unique even if they load the same article because publishers frequently include session and navigational parameters into the URL.[2] We therefore "stem" all URLs into a canonical format that removes most query parameters except for certain cases such as `?id=` which are sometimes used to index pages.

A-1.4 Landing and Article Pages

It is important for us to distinguish between article pages and landing pages (such as `http://www.nytimes.com` or `http://www.nytimes.com/business`). For each canonical URL, we count the number of page loads on every single day across all users and order days in reverse order according to page count. We then analyze how many days it takes to generate 80% of all page loads – if it takes more than 15 days then we regard a URL as landing page and otherwise an article. Intuitively, we exploit the fact that articles tend to appear only a small number of days and therefore generate the bulk of page views within that time frame. The NYT homepage, on the other hand, is loaded at a fairly constant rate on every single day.

When evaluated by human auditing, we found that this simple algorithm successfully classifies about 95% of all canonical URLs correctly.

[2]Such as `https://www.nytimes.com?origin=google`, for example.

A-1.5 News Minisessions and Referral Modes

Most users spend only a fraction of time browsing news during a browsing session - the remainder is spent on Facebook, email etc. Thanks to the referrer URLs, we can think of news browsing within a browsing session as a set of "trees": each tree has a single root (a URL without a news referrer) and a number of branches that each have a (news) referrer URL that indirectly links to the root.

We call such a tree a "news mini session" (or NMS). We assign a referral mode to each NMS depending on whether the root page load was the result of direct navigation (such as a bookmark), a search on Google, Bing or Yahoo, navigation from Twitter or Facebook (such as clicking a news article on these platforms), a referral from Google News or some other origin (such as clicking on an email).

A-2 Topic Classification

We create topics for news articles by mapping unique news article URLs (excluding landing pages) to Spanish language Wikipedia articles. We exploit the fact that Wikipedia covers most major news events within a short period of time. This provides us with a stable taxonomy for classifying a corpus of news articles. After automatically generating the clusters of articles, we manually name topics by their "super-topics" (high-level topics) and the "sub-topics" that in general correspond to the specific subject of matched Wikipedia articles.

A-2.1 Data

Scraping. We scrape all unique URLs and extract the text using Boilerpipe which is an open-source program that can detect text within an html page (optimized for newspaper articles). Within each publisher we run a cleaning algorithm which compares the article text of all articles and identifies repeated sentences across articles. This algorithm removes boilerplate text such as privacy statements which otherwise pollute the article text. Sentences that occur in distinct URLs on more than 3 days a week over a one month period are removed using this heuristic.

Wikipedia. Wikimedia holds an up-to-date online repository of all Wikipedia articles for several languages. We extract all Wikipedia articles with more than 100 words. We also extract outlinks (connections between Wikipedia articles).

A-2.2 Pre-processing

We build a dictionary of stemmed words using a standard Spanish language stemmer from all Wikipedia concepts (a concept is really an article on Wikipedia - we refer to Wikipedia articles as "concepts" in order to avoid confusion). We ignore rare words (words that appear

in fewer than 3 wikipedia articles) and Spanish stop words (such as articles and frequent words such as the Spanish equivalents of "and" and "or"). We calculate the IDF (inverse document frequency) score for each dictionary word defined as:

$$\log\left(\frac{\#\text{ all words})}{\#\text{ Wikipedia concepts in which specific word appears}}\right)$$

We do the same for each bigram on Wikipedia.

For each newspaper article we find all words and bigrams. We count within each newspaper article (including title) the number of occurrences of each word and bigram and calculate the TFIDF score as the number of occurrences times IDF score.[3] We then order all words and bigrams within the article according to this score in reverse order and take the top 30 words and top 30 bigrams. For each article, we call those top 30 keywords and top 30 bigrams the *article fingerprint*.

A-2.3 Newspaper Article to Wikipedia Concept matching

In this step we match Wikipedia concepts to each newspaper article.

We calculate a word-based newspaper article/Wikipedia concept similarity score by taking the top 25 keywords of the article (out of 30). We then iterate through all Wikipedia concepts. For each Wikipedia concept we (i) check which of these 25 words appear in the Wikipedia concept; (ii) construct the sum of TFIDF scores for those words; (iii) divide this sum by the sum of all 25 TFIDF scores. This provides us a score between 0 and 1 that captures the similarity between the newspaper article and the Wikipedia concept based on words. We calculate an analogous bigram-based similarity score.

Each of these two scores lies between 0 and 1. We calculate the weighted average for each news article/Wikipedia article pair by placing weight 0.6 on the word score. We create a list of the top 10 Wikipedia concepts according to this weighted similarity score.[4]

A-2.4 Pure Macro Topics

Many Wikipedia concepts refer to very similar ideas. For example, the English language Wikipedia has several articles related to the passage of the Affordable Care Act. In order to define clean topics it is necessary to cluster closely related concepts into what we call a "pure topic." In order to accomplish this clustering we construct a graph that connects Wikipedia concepts. This graph is based on two subgraphs – the *co-occurence graph* and the *common outlink graph*.

[3] TFIDF scores are a standard scoring rule that is commonly used for information retrieval systems.

[4] We determined this weight as well as the number of keywords and bigrams on which the weight is based by hiring student raters who evaluated the resulting Wikipedia article assignment for a sample of newspaper articles. We then determined the three parameters through a grid search that maximized the overall evaluation score.

The co-occurence graph is constructed by assigning an edge weight to each pair of Wikipedia concepts i and j as follows: we find the number of newspaper articles that (i) list both these concepts among the top 3 matches and (ii) assign scores to both concepts of at least 0.5. Intuitively, the co-occurrence graph assigns a large weight to pairs of substitutable concepts.

For the outlink graph the weight between two concepts i and j is defined as follows: we count the number of concepts that both i and j either link to or are linked from (common neighbors). We then normalize this score by dividing by the size of the union of all articles that link to or are linked from i *or* j.

Finally, we merge both graphs by multiplying the edge weights (where some weights might go to zero as a result of this multiplication). Intuitively, Wikipedia concepts that are close in this merged graph tend to both have a high co-occurrence count and a high share of common outlinks. We use the fast Louvain algorithm (which maximizes modularity) to detect clusters (Blondel et al., 2008). We call these resulting clusters the "pure topics."

Going back to the articles, every newspaper article whose top-scoring Wikipedia concept has at least weighted score 0.5 and is part of such a pure topic is then assigned to this pure topic. Articles whose top-scoring concept has a score of less than 0.5 or that do not connect to any concept in a pure topic are left unassigned at this stage.

A-2.5 Topic Augmentation

In this step, we assign some of the so far unassigned newspaper articles to pure topics. We again use a clustering algorithm which we first overview before providing details. The basic idea is to construct a super-network of all articles (both assigned and unassigned). We create strong links between assigned articles in this super-network that belong to the same pure macro topic. This ensures that these articles will be assigned to the same clusters when finding communities in the super-network. We also add links between unassigned and assigned articles based on semantic and word similarity.

In an ideal world, we would construct the super-network using all articles. The problem is that with N million articles, calculating edge weights between all pairs of articles requires N^2 calculations which is infeasible given the number of unique articles. To reduce the computation time, we use a short-cut that iteratively adds nodes to the network and groups them into communities.

We take the set of pure topics such that the page view count makes up 90% of assigned articles. For each pure topic we take 20 articles which have a top weighted Wikipedia concept score above 0.9, and we sample by page view weight (these tend to be representative articles that clearly belong to the respective pure macro topic). We construct a base network of "islands" where each island is a complete graph that connects all 20 base article belonging to a particular macro topic (weight 1 on each edge). Hence, there are as many islands as pure topics. The total number of base articles M is equal to 20 times the number of pure topics.

We then take batches of M/2 unassigned articles and embed them in this base network. We connect them to each of the M base articles by using a weighted average of the following two sub-weight.

1. The first sub-weight w_{ij} between an unassigned article i and base article j is determined by semantic similarity (distance between the associated Wikipedia concepts). We take the top 5 associated Wikipedia concepts for both i and j, and then take the intersection of both concepts sets. Then we sum i's scores for these overlapping concepts and divide by the sum of i's top 5 scores to get a number between 0 and 1. Analogously, we calculate the corresponding score for j. The average of these two numbers defines our first sub-weight w_{ij}. Intuitively, this weight is large if both newspaper articles are associated with similar Wikipedia concepts and more important concepts receive higher weight. We then calculate the top 1 percentile cutoff over all these sub-weights and set all the weights below this cutoff to 0. This "pruning" reduces noise.

2. The second sub-weight $\tilde{w}_{ij}$ measures word and bigram similarity between the two articles. We consider word similarity first. We take the top 8 keywords for articles i and j and the associated TFIDF scores from both fingerprints. We then use the same procedure as for semantic similarity. We then repeat the procedure again for the top 5 bigrams. We average the resulting scores by putting weight 0.6 on the word score. This provides us with the second sub-weight $\tilde{w}_{ij}$. Intuitively, this weight is large if both articles are associated with similar keywords and bigrams. We use the same 1 percentile pruning as applied to the semantic network.

We add both sub-weights and divide by half. This gives us a $[0, 1]$ weight that defines the graph links between the base network and the unassigned batch articles. We again run the Louvain community detection on each resulting super-network, such that we always preserve the original pure macro topics but "augment" them with the previously unassigned articles.

The algorithm scales linearly with the dataset.

A-2.6 Supervised cleaning

The last step is the only supervised step that requires manual intervention. Recall, that each macro topic is defined by the set of associated Wikipedia concepts. We have student evaluators define English-language nested labels for these 320 labels such as *Health: Spain Ebola crisis* for the macro topic that describes the 2015 Ebola crisis in Spain. Our nesting has two levels: super-topic (*Health*) and sub-topic (*Spain Ebola crisis*). This makes it easy to evaluate the final topic assignment by separately rating the super-topic and sub-topic assignments.